

Curve Fitting in Nuclear Magnetic Resonance Spectroscopy: Illustrative Examples Using a Spreadsheet and Microcomputer

Peter L. Irwin,* William C. Damert, and Landis W. Doner

*United States Department of Agriculture
Agricultural Research Service
Eastern Regional Research Center
600 E. Mermaid Lane
Philadelphia, Pennsylvania 19118*

Received June 6, 1993; Revised September 13, 1993

A simplified explanation of the modified Gauss-Newton linearization (curve-fitting) procedure is presented by translating this algorithm into a spreadsheet with examples from various single- and multiple-pulse NMR experiments. Comparisons are made with linear regression analysis of mathematically manipulated data.

INTRODUCTION

Numerical methods used to fit various nonlinear functions to data have been available for more than 30 years (1). However, for nonlinear parameter estimation many people continue to use techniques that rely on linear regression analysis (2). The problem is that direct linearization sometimes provides a good model only on a restricted domain (3, 4). Thus it is often advantageous to be able to calculate a solution to a nonlinear expression, or model, directly.

In this article we review the modified Gauss-Newton (1, 5) method, one of the most common techniques used for curve fitting in specialized statistical programs (6) or software packages associated with NMR spectrometers. To illustrate the numerical formalism and ease of operation of this method, we created a modified Gauss-Newton (G-N) template from common spreadsheet software. Only recently has it been possible to perform these calculations on microcomputers through the manipulation of relatively large matrices. Except for speed, the template we describe herein works as well as any of the equivalent curve-fitting programs on larger computers with which we are familiar. All that is numerically required for a G-N template is the ability to do matrix transposition, multiplication, and inversion, and to solve recurrence relations (equations involving "circular references;" e.g., formulas and/or cells that refer to themselves for solution) (7). Problems involving circular references require repeated calculations, or iterations, until a specific numeric condition—usually some minimal change as defined by the numeric precision—is met.

* To whom correspondence should be addressed. Reference to a brand or firm name does not constitute endorsement by the U.S. Department of Agriculture over others of a similar nature not mentioned.

G-N LINEARIZATION

The most commonly used curve-fitting techniques (nonlinear least-squares approximation) are Gauss-Newton linearization (G-N), steepest descent, and the Levenberg-Marquardt method (5, 6). In the G-N method (1) a Taylor series is used to linearize the function of interest. Say we have a function, F , that depends on parameters, P_l $\{l = 1, 2, \dots, L\}$, which we want to fit to observations, Y_m , that were measured at different values of X_m $\{m = 1, 2, \dots, M\}$. For simplicity's sake, let us make the total number of parameters small (e.g., $L = 2$). The difference between Y_m and F , calculated with parameter values P_1 and P_2 , is

$$\Delta Y_m = Y_m - F(P_1, P_2; X_m) \quad [1]$$

Assuming that the errors in the measured Y values are all equal, we seek the values of P_1 and P_2 , which minimize the sum of the squares of these deviations. At any n th stage of iteration the parameter values are P_1^n and P_2^n , and the value of F is

$$F_m^n = F(P_1^n, P_2^n; X_m) \quad [2]$$

For (P_1, P_2) near (P_1^n, P_2^n) one can expand F_m in a Taylor series around (P_1^n, P_2^n) . Retaining only first-order terms yields the expression

$$\Delta Y_m \cong Y_m - F_m^n - \frac{\partial F_m^n}{\partial P_1} (P_1 - P_1^n) - \frac{\partial F_m^n}{\partial P_2} (P_2 - P_2^n) \quad [3]$$

$$= Y_m - F_m^n - \sum_{l=1}^2 \frac{\partial F_m^n}{\partial P_l} (P_l - P_l^n) \quad [4]$$

With the above approximation, the error sum of squares is

$$\sigma(P_1, P_2) \cong \sum_{m=1}^M (\Delta Y_m)^2 \quad [5]$$

$$\cong \sum_{m=1}^M \left[Y_m - F_m^n - \sum_{l=1}^2 \frac{\partial F_m^n}{\partial P_l} (P_l - P_l^n) \right]^2 \quad [6]$$

To minimize, the derivatives (gradient) of σ are set to zero such that ($\lambda = 1$ or 2)

$$\frac{\partial \sigma}{\partial P_\lambda} = -\sum_{m=1}^M 2 \left[Y_m - F_m^n - \sum_{l=1}^2 \frac{\partial F_m^n}{\partial P_l} (P_l - P_l^n) \right] \frac{\partial F_m^n}{\partial P_\lambda} = 0 \quad [7]$$

One may simplify the algebra by introducing vector and matrix notation (8, 9). In this notation, \mathbf{Z} ($M \times 2$) represents the gradient of F^n evaluated at X_1, \dots, X_{M-1}, X_M and contains elements $Z_{ml} \equiv \partial F_m^n / \partial P_l$. \mathbf{Z}' , a $2 \times M$ matrix, is the transpose of \mathbf{Z} . Similarly, the vector \mathbf{P} (2×1) contains elements P_l . Applying this notation and rearranging Eq. [7] gives

$$\mathbf{Z}'(\mathbf{Y} - \mathbf{F}^n) = \mathbf{Z}'\mathbf{Z}(\mathbf{P} - \mathbf{P}^n) \quad [8]$$

which, when written out completely, gives

$$\begin{bmatrix} \frac{\partial F_1^n}{\partial P_1} & \frac{\partial F_2^n}{\partial P_1} & \cdots & \frac{\partial F_M^n}{\partial P_1} \\ \frac{\partial F_1^n}{\partial P_2} & \frac{\partial F_2^n}{\partial P_2} & \cdots & \frac{\partial F_M^n}{\partial P_2} \end{bmatrix} \begin{bmatrix} Y_1 - F_1^n \\ Y_2 - F_2^n \\ \cdots \\ Y_M - F_M^n \end{bmatrix} = \begin{bmatrix} \frac{\partial F_1^n}{\partial P_1} & \frac{\partial F_2^n}{\partial P_1} & \cdots & \frac{\partial F_M^n}{\partial P_1} \\ \frac{\partial F_1^n}{\partial P_2} & \frac{\partial F_2^n}{\partial P_2} & \cdots & \frac{\partial F_M^n}{\partial P_2} \end{bmatrix} \begin{bmatrix} \frac{\partial F_1^n}{\partial P_1} & \frac{\partial F_1^n}{\partial P_2} \\ \frac{\partial F_2^n}{\partial P_1} & \frac{\partial F_2^n}{\partial P_2} \\ \cdots & \cdots \\ \frac{\partial F_M^n}{\partial P_1} & \frac{\partial F_M^n}{\partial P_2} \end{bmatrix} \begin{bmatrix} P_1 - P_1^n \\ P_2 - P_2^n \end{bmatrix} \quad [9]$$

Our goal is to solve for $\mathbf{P} - \mathbf{P}^n$ and, because $\mathbf{Z}'\mathbf{Z}$ is a square (2×2) matrix, each side of the equation is multiplied by its inverse, $(\mathbf{Z}'\mathbf{Z})^{-1}$ assuming that the matrix is nonsingular (we will mention the singular case later). This manipulation gives the expression for the parameter change at the n th iteration

$$\Delta \mathbf{P} = (\mathbf{P} - \mathbf{P}^n) = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'(\mathbf{Y} - \mathbf{F}^n) \quad [10]$$

The term $\Delta \mathbf{P}$ is considered to have converged ($n = N$) when values from the previous calculation, P_i , approximately equal values from the latest iteration, P_i^n .

By applying the theory of linear regression to the approximate linear equation we obtain an expression for the standard error in the parameter estimates. For the l th parameter, the standard error is given (5) by

$$(SE)_l = \sqrt{\frac{(\mathbf{Z}'\mathbf{Z})_{ll}^{-1} \sum_{m=1}^M [Y_m - F_m^N]^2}{M - L}} \quad [11]$$

where $(\mathbf{Z}'\mathbf{Z})_{ll}^{-1}$ is the ll th element of the square $(L \times L)$ $(\mathbf{Z}'\mathbf{Z})^{-1}$ matrix [e.g., the standard error for P_2^N is equal to the above equation with $(\mathbf{Z}'\mathbf{Z})_{ll}^{-1} = (\mathbf{Z}'\mathbf{Z})_{22}^{-1}$]. This term, which is used as an estimate of the standard error in the final estimate of P_i (P_i^N), is often referred to as the asymptotic standard error because it is derived only from the approximate linear equation. Using this estimate, one would quote the estimate of P_i as

$$P_i = P_i^N \pm (SE)_i \quad [12]$$

The G-N method is widely used because, in theory at least, it should always converge. Of course, problems can be found in any curve-fitting method. However, when we have used the G-N linearization method (Table 1) we have not experienced the problems with nonconvergence or overshooting that have been reported elsewhere (5). Although we have found the methods described here to be suitable for a large variety of problems, certain pitfalls associated with any method used to solve nonlinear least squares problems should be noted.

First, unlike the linear case, there is no guarantee of a unique solution to a nonlinear problem. Almost all methods for obtaining solutions to nonlinear problems, including the one outlined above, rely on iterating from an initial set of parameter estimates. The fact that these iterations converge to a certain final value does not necessarily imply that another solution with an even smaller sum of squares of residuals might not exist and be reachable from another set of initial estimates. In practice, for problems such as those considered here, we believe that the existence of more than one physically reasonable solution is highly unlikely.

TABLE 1
Common NMR-Related Functions We Have Tested with the G-N Spreadsheet

Parameter(s) of Interest	Purpose of Experiment	Function to Fit
$T_{CH}, T_{1\rho}, I_0$	Contact time study*	$\frac{I_0}{1 - \frac{T_{CH}}{T_{1\rho}}} \left[1 - e^{\frac{T_{CH}}{T_{1\rho}} - 1 \left(\frac{X_m - \Delta\tau}{T_{CH}} \right)} \right] e^{-\left(\frac{X_m - \Delta\tau}{T_{1\rho}} \right)}; X_m = \tau \quad [13]$
T_1	Inversion recovery	$I_\infty \left[1 - 2e^{\frac{-[X_m - \Delta\tau]}{T_1}} \right]; X_m = \tau \quad [14]$
T_2	Carr-Purcell-Meiboom-Gill	$I_0 e^{-\frac{X_m}{T_2}} + \gamma; X_m = 2\tau \quad [15]$
PW_{180°	Probe performance*	$I_{\max} = \sin \left[\frac{X_m \pi}{PW_{180^\circ}} \right]; X_m = PW_m \quad [16]$
σ_i, I_i, δ_i	Gaussian deconvolution*	$\sum_i I_i e^{-\frac{(\delta_i - X_m)^2}{2\sigma_i^2}}; X_m = \delta_{\text{observed}} \quad [17]$
σ_i, I_i, δ_i	Lorentzian deconvolution*	$\sum_i \frac{I_i \sigma_i^2}{(X_m - \delta_i)^2 + \sigma_i^2}; X_m = \delta_{\text{observed}} \quad [18]$
κ	Spin-exchange processes	$I_{\max} \left[1 - e^{(X_m - \kappa)} \right]; X_m = \tau_{\text{mixing}} \quad [19]$
* Function is not readily transformable for linear regression analysis		

Second, one should not attempt to obtain parameters with inadequate data or an inappropriate model. A trivial example would be attempting to find the parameters A and B in a model

$$Y = AX + B\sin(X) \quad [20]$$

where data exist only for small X_m . Because $\sin(X_m) \cong X_m$ for values of X_m near $X_m = 0$, the equation may be approximated by

$$Y \cong (A + B)X \quad [21]$$

which indicates that we can determine the sum, $A + B$, but not A or B individually. This ill-posedness manifests itself as $Z'Z$ becomes almost singular when the values of X are all close to zero. In a similar fashion, the matrix for solving the inversion recovery equation (Eq. [14]), would be nearly singular if one attempted to solve for I_∞ by using only data for small values of τ . In these and less obvious cases, this ill-conditioning will often be apparent by the large standard errors for the parameters.

G-N LINEARIZATION ON A SPREADSHEET

To demonstrate the G-N template (Fig. 1A and 1B; Macintosh IIfx with Excel 4.0 and System 7.0 with 8 Mb RAM) we have chosen the four-parameter equation,

Curve Fitting in Nuclear Magnetic Resonance Spectroscopy

$$F_m = F(P_1 = I_0, P_2 = T_{CH}, P_3 = T_{1\rho}, P_4 = \Delta\tau; X_m = \tau) \quad [22]$$

$$= \frac{I_0}{1 - \frac{T_{CH}}{T_{1\rho}}} \left[1 - e^{\frac{T_{CH}}{T_{1\rho}} - 1 \left(\frac{\tau - \Delta\tau}{T_{CH}} \right)} \right] e^{-\left(\frac{\tau - \Delta\tau}{T_{1\rho}} \right)} \quad [23]$$

which describes the buildup and eventual loss of ^{13}C magnetization caused by polarization transfer from remotely or directly attached ^1H s (I_0). The data to be fit are linear baseline-corrected and integrated $^{13}\text{C}=\text{O}$ resonances, $Y_m = I_{C=\text{O}}$, which change as a function of different polarization transfer, or contact, times ($X_m = \tau$).

	A	B	C	D	E	F	G	H	I	J
1	X_m	Y_m	I_0	$T_{CH} / \mu\text{s}$	$T_{1\rho} / \mu\text{s}$	$\Delta\tau / \mu\text{s}$	$\text{Error SS} = \sum_{m=1}^{M=10} [Y_m - F_m]^2$			
2	25	0.5	13.96	853.16	2053.40	-27.62	0.391745186			
3	50	1.51	I_0'	T_{CH}'	$T_{1\rho}'$	$\Delta\tau'$				
4	100	2.01	13.96	853.16	2053.40	-27.62				
5	200	2.95								
6	400	4.71								
7	800	6.79								
8	1000	7.54								
9	2000	6.6								
10	4000	3.02								
11	6000	1.38								
12			$(Z'Z)^{-1}Z'[Y_m - F_m]$	$Y_m - F_m$			$\frac{\partial F_m}{\partial I_0}$	$\frac{\partial F_m}{\partial T_{CH}}$	$\frac{\partial F_m}{\partial T_{1\rho}}$	$\frac{\partial F_m}{\partial \Delta\tau}$
13			-0.037151446	-0.324274421	$m=1$		0.059045446	-0.000936528	5.17425E-06	-0.014982576
14			2.409951017	0.318989769	$m=2$		0.085315919	-0.001333058	1.10597E-05	-0.014359723
15			-5.19530465	0.130967946	$m=3$		0.13460115	-0.002040112	2.88508E-05	-0.013174267
16			0.947299106	-0.13674248	$m=4$		0.221113358	-0.003147914	8.54816E-05	-0.011027794
17				-0.214995283	$m=5$		0.352793358	-0.004396508	0.000261916	-0.007513929
18				-0.117495272	$m=6$		0.494806252	-0.004538506	0.000741638	-0.002838492
19				0.221901117	$m=7$		0.524219118	-0.00401302	0.000995558	-0.001342393
20				-0.0791843	$m=8$		0.478451597	-0.000613472	0.001966338	0.001733209
21				-0.126551395	$m=9$		0.225397664	0.001444403	0.002119613	0.001386606
22				0.1320306	$m=10$		0.089396089	0.000870776	0.001381196	0.000593776
23										
24										
25										
26										
27										
28										

	A	B	C	D	E	F	G	H	I	J
29	B									
30										
31										
32	$Z' = \text{TRANSPOSE}(F10:I19) (4 \times 10 \text{ M})$									
33	0.059045446	0.085315919	0.13460115	0.221113358	0.352793358	0.494806252	0.524219118	0.478451597	0.225397664	0.089396089
34	-0.000936528	-0.001333058	-0.002040112	-0.003147914	-0.004396508	-0.004538506	-0.00401302	-0.000613472	0.001444403	0.000870776
35	5.17425E-06	1.10597E-05	2.88508E-05	8.54816E-05	0.000261916	0.000741638	0.000995558	0.001966338	0.002119613	0.001381196
36	-0.014982576	-0.014359723	-0.013174267	-0.011027794	-0.007513929	-0.002838492	-0.001342393	0.001733209	0.001386606	0.000593776
37										
38										
39										
40										
41	$Z'Z = \text{MMULT}(A33:J36;F10:I19) (4 \times 4 \text{ L})$									
42	1.00958752	-0.006930226	0.00254732	-0.009885631						
43	-0.006930226	7.59781E-05	-5.80218E-06	0.000147527						
44	0.00254732	-5.80218E-06	1.1885E-05	1.98582E-07						
45	-0.009885631	0.000147527	1.98582E-07	0.00079745						
46										
47										
48										
49										
50	$(Z'Z)^{-1} = \text{MINVERSE}(A42:D45) (4 \times 4 \text{ L})$									
51	59.83161857	5025.94737	-10367.01919	-185.5029149						
52	5025.94737	444013.0336	-860122.7394	-19622.92294						
53	-10367.01919	-860122.7394	1885700.443	30136.129						
54	-185.5029149	-19622.92294	30136.129	2577.097408						
55										
56										
57										
58										
59										
60	$(Z'Z)^{-1}Z' = \text{MMULT}(A52:D55;A33:J36) (4 \times 10 \text{ M})$									
61										
62	1.551515844	0.953824677	-0.05532433	-1.432182849	-2.309854533	-0.367258531	1.123701054	4.836672209	-1.485887288	-4.703844944
63	170.4802598	109.1652437	4.364234885	-143.5379818	-258.8208214	-110.4857011	22.90419833	408.9799777	-78.16600917	-363.7137274
64	-248.3562937	-149.7894791	16.71684631	244.1638989	391.5757084	86.97193043	-146.0304959	-672.29828	459.6763022	946.6706907
65	-31.03131752	-26.34096559	-18.01785881	-5.089508589	9.357154891	12.30577509	8.045792515	-12.99157499	-2.704986555	9.483713377

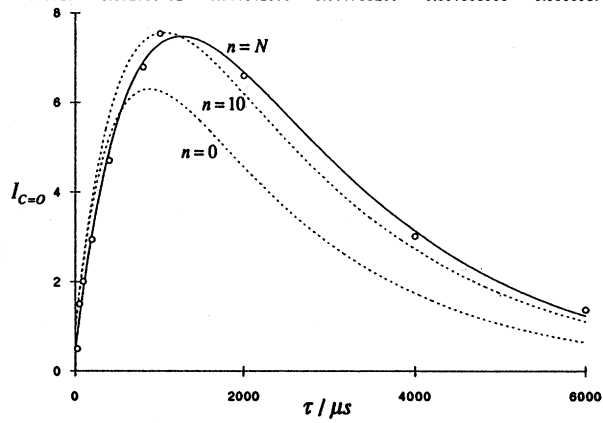


Figure 1. G-N linearization spreadsheet for fitting CPMAS NMR contact time data to Eq. [13]. Part (A) contains initial (P_i) and final (P_f) values for the parameters I_0 , T_{CH} , $T_{1\rho}$, and $\Delta\tau$ as well as Z , ΔP , and standard error arrays. Part (B) contains all the transformations of Z used in calculating ΔP as well as the plot of data and curve fits at $n = 0$, 10, and 200 iterations. For illustrative purposes, the convergence rate factors were made exceptionally small ($R_i = 0.01 - 0.05$).

The spreadsheet (Fig. 1) was created by carrying out the following steps.

1. X_m {A2:A11} and Y_m {B2:B11} data were input (Fig. 1A).
2. Initial guesses ($n = 0$) for parameters (P_i) were keyed into cells C2 ($I_0 = 10$), D2 ($T_{CH} = 500$ μ s), E2 ($T_{1\rho} = 2,000$ μ s), and F2 ($\Delta\tau = -50$ μ s); I_0 is the maximum of the C=O magnetization if no rotating-frame spin-lattice relaxation occurs; T_{CH} is the reciprocal exponential rate constant describing the buildup of C=O magnetization; $T_{1\rho}$ is the rotating-frame spin-lattice relaxation time describing the loss of C=O magnetization as a function of spin-lock time or ^1H - ^{13}C polarization transfer time; and $\Delta\tau$ is the time translation correction that is needed because $I > 0$ when $\tau = 0$.

3. The $Y_m - F_m^n$ array ($m = 1-10$) was created such that

$$\begin{aligned} \{D10:D19\} = & \{B2:B11\} - ((\$C\$2/(1 - (\$D\$2/\$E\$2)) * (1 - \text{EXP}(((\$D\$2/\$E\$2) - 1) * \\ & ((\{A2:A11\} - (\$F\$2))/\$D\$2))) * (\text{EXP}(-(\{A2:A11\} - (\$F\$2))/\$E\$2)))) \end{aligned} \quad [24]$$

Equalities within brackets {} indicate that the expression is applied throughout a block of cells. Cells defined as \$LETTER\$NUMBER indicate that each cell within the array contains this particular, or absolute, reference regardless of its position within the array; for example, when $m = 8$

$$\begin{aligned} D17 = & B9 - ((\$C\$2/(1 - (\$D\$2/\$E\$2)) * (1 - \text{EXP}(((\$D\$2/\$E\$2) - 1) * ((A9) - \\ & (\$F\$2))/\$D\$2))) * (\text{EXP}(-((A9) - (\$F\$2))/\$E\$2)))) = -0.0791843 \end{aligned} \quad [25]$$

4. The error sum of squares estimation, $\sum_{m=1}^{M=10} [Y_m - F_m^n]^2$, was calculated whereupon

$$G4 = \{\text{SUM}((D10:D19)^2)\} \quad [26]$$

5. A collection of derivative expressions for the evaluation of Z was created

$$Z = \{F10:I19\} = \begin{bmatrix} \frac{\partial F_1^n}{\partial I_0} & \frac{\partial F_1^n}{\partial T_{CH}} & \frac{\partial F_1^n}{\partial T_{1\rho}} & \frac{\partial F_1^n}{\partial \Delta\tau} \\ \dots & \dots & \dots & \dots \\ \frac{\partial F_{M-1}^n}{\partial I_0} & \frac{\partial F_{M-1}^n}{\partial T_{CH}} & \frac{\partial F_{M-1}^n}{\partial T_{1\rho}} & \frac{\partial F_{M-1}^n}{\partial \Delta\tau} \\ \frac{\partial F_M^n}{\partial I_0} & \frac{\partial F_M^n}{\partial T_{CH}} & \frac{\partial F_M^n}{\partial T_{1\rho}} & \frac{\partial F_M^n}{\partial \Delta\tau} \end{bmatrix} \quad [27]$$

6. Z' , {A33:J36}, was calculated by selecting the appropriate ($4 \times 10 = L \times M$) cells and performing the function {=TRANSPOSE(F10:I19)} (Fig. 1B).
7. $Z'Z$, {A42:D45} ($L \times L$), {=MMULT(A33:J36,F10:I19)}
8. $Z'Z^{-1}$, {A52:D55} ($L \times L$), {=MINVERSE(A42:D45)}
9. $Z'Z^{-1}Z'$, {A62:J65} ($L \times M$), {=MMULT(A52:D55,A33:J36)}
10. ΔP , {C10:C13} ($4 \times 10 \times 10 \times 1 = 4 \times 1$), {=MMULT(A62:J65,D10:D19)}
11. P_i^n values, {C4:F4}, were calculated as

$$P_1^n = I_0^n = C4 = C2 + (C10 * R_1) \quad [28]$$

$$P_2^n = T_{CH}^n = D4 = D2 + (C11 * R_2) \quad [29]$$

$$P_3^n = T_{1\rho}^n = E4 = E2 + (C12 \cdot R_3) \quad [30]$$

$$P_4^n = \Delta\tau^n = F4 = F2 + (C13 \cdot R_4) \quad [31]$$

Multiplying ΔP_i ($\Delta P_1 = C10$, $\Delta P_2 = C11$, etc.) by a fraction (R_i , the convergence rate factor for P_i^n) allows one to slow down the convergence process so that the initial guesses do not have to be as good, and potential problems with overshooting are avoided.

12. Cells {C2:F2} ($P_{1:4}$) were made equal to cells {C4:F4} ($P_{1:4}^N$) to create the circular reference needed to induce iterative solving; the precision desired can be changed by selecting the appropriate P_i cell (C2 to F2) and changing the number of decimal places.

A plot of the data as well as the initial ($n = 0$), intermediate ($n = 10$), and final ($N = 200$) fits are shown in Fig. 1B. In Fig. 2 we present the change in P_i^n and the error sum of squares at various stages of iteration to demonstrate graphically what occurs during the convergence process. Notice that even though $T_{1\rho}$ initially increases rapidly as a function of n , it nevertheless reverses direction and converges at a value only somewhat higher than the initial guess; of course, the error sum of squares asymptotically approaches zero as n approaches N . One of the greatest criticisms of this linearization method is that P_i^n values sometimes oscillate around P_i^N and slow the convergence process (5). In this example (Fig. 1) and others (Table 1) VAX 8350 equivalent calculations were about five times faster (e.g., <1 min) and resulted in similar estimates of P_i^N . When we performed these same calculations on a smaller microcomputer we obtained identical results except for a somewhat slower computational time.

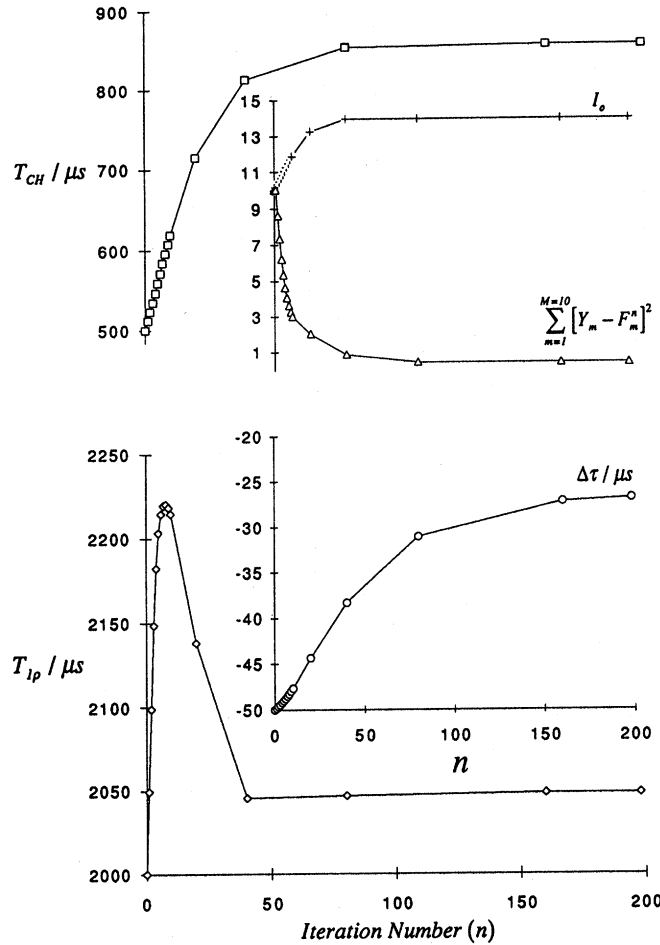


Figure 2. Plots of I_0 , T_{CH} , $T_{1\rho}$, $\Delta\tau$, and $\sum_{m=1}^M [Y_m - F_m^n]^2$ derived from the contact time fits at various iterations, n .

With any template as a starting point, about 30 min is required for modification to a new function. Much of this time, which is needed to make the derivative matrix, \mathbf{Z} , can be eliminated by using software packages (11) that symbolically solve partial derivatives quickly (<30 s). Certain software packages do not use functional values of derivatives; instead, they compute the numerical equivalent at each stage of iteration.

G-N LINEARIZATION AND LINEAR REGRESSION

Linear regression analysis of nonlinear expressions sometimes can disguise deviations from the fit because they involve the mathematical transformation of the original \mathbf{Y} and/or \mathbf{X} vectors. Figures 3 and 4 show results of traditional T_1 (2, 12) and T_2 (13) experiments, respectively, in both the nonlinear G-N (A) and log-linear (B) forms. The inversion recovery data (Fig. 3A, $T_1 = 313 \pm 0.007$ ms) show an excellent fit with the G-N method, described above, because the deviation from the fit (Fig. 3C)

$$\frac{Y_m - F_m^N}{Y_m} \quad [32]$$

averaged only about $\pm 1\%$. The linear regression method, shown here in semilog format (Fig. 3B, $T_1 = 297 \pm 4.89$ ms), also displays a good fit. However, only 50-70% of the original τ values are usable; the deviation from the retransformed semilog fit (Fig. 3D) was about threefold greater

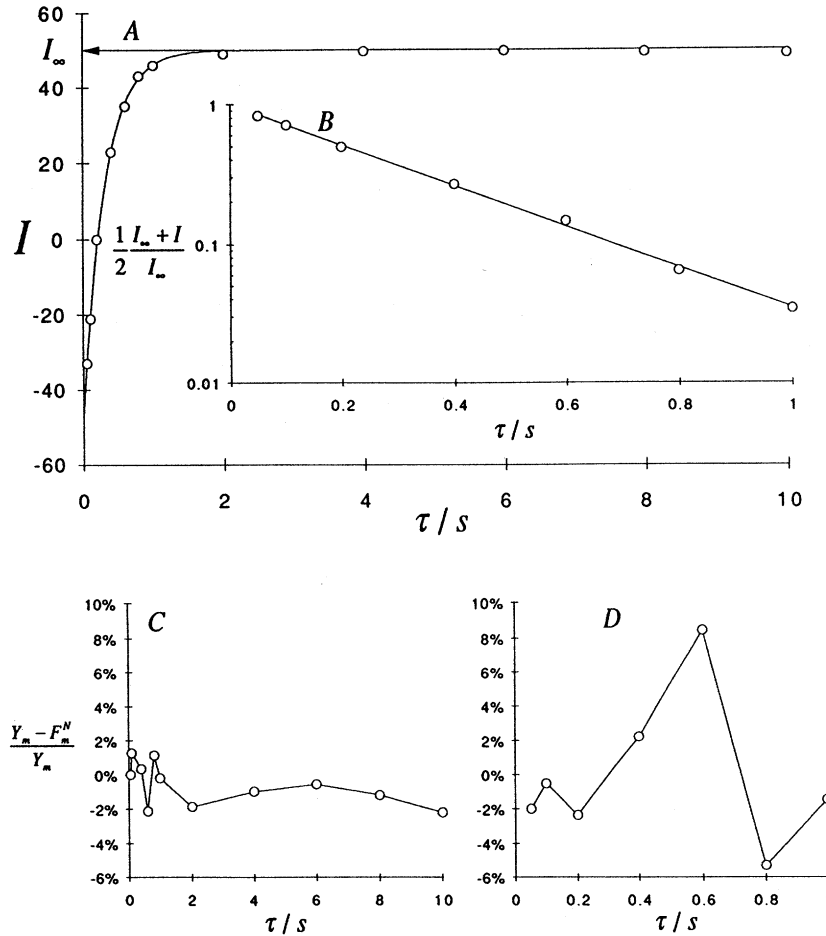


Figure 3. Inversion recovery (Eq. [14]) data plotted using the G-N fit (A) and the semilog plot (B). Deviations in $(Y_m - F_m^N)/Y_m$ as a function of τ for (A) and (B) are shown in (C) and (D), respectively.

than that of the G-N method. Statistically speaking, a comparison of residuals derived from the G-N method and the log-linear model is not valid, but it does give a sense of the goodness of the two fits. The major statistical criticism of the linear regression technique for T_1 is that least-squares analysis on transformed data (Y^*) can fail because it is based on the incorrect error model. For example, nonlinear regression minimizes the sum of squares of the deviations between the measured and the calculated Y vectors, as shown in Eq. [5]. To use linear regression on Eq. [14] one must first transform the equation to get it into linear form,

$$Y_m^* = \log \left[\frac{1}{2} \left(1 + \frac{Y_m}{I_\infty} \right) \right] = -\frac{X_m}{T_1} + b \quad [33]$$

which requires that

$$Y_m^* = y_m^* + \epsilon_m^* \quad [34]$$

where the transformed error term, ϵ_m^* , is independent and identically distributed normal (i.i.d. normal). Linear regression can be used to find the solution that minimizes the sum of squares of the measured and calculated values of Y^* and is simply a different problem than that of minimizing $\sum_{m=1}^M [Y_m - F_m^n]^2$ directly. Another problem with the linear regression technique for T_1 is that one must know I as τ approaches ∞ (I_∞). Thus several $I(\tau)$ values should be measured by at

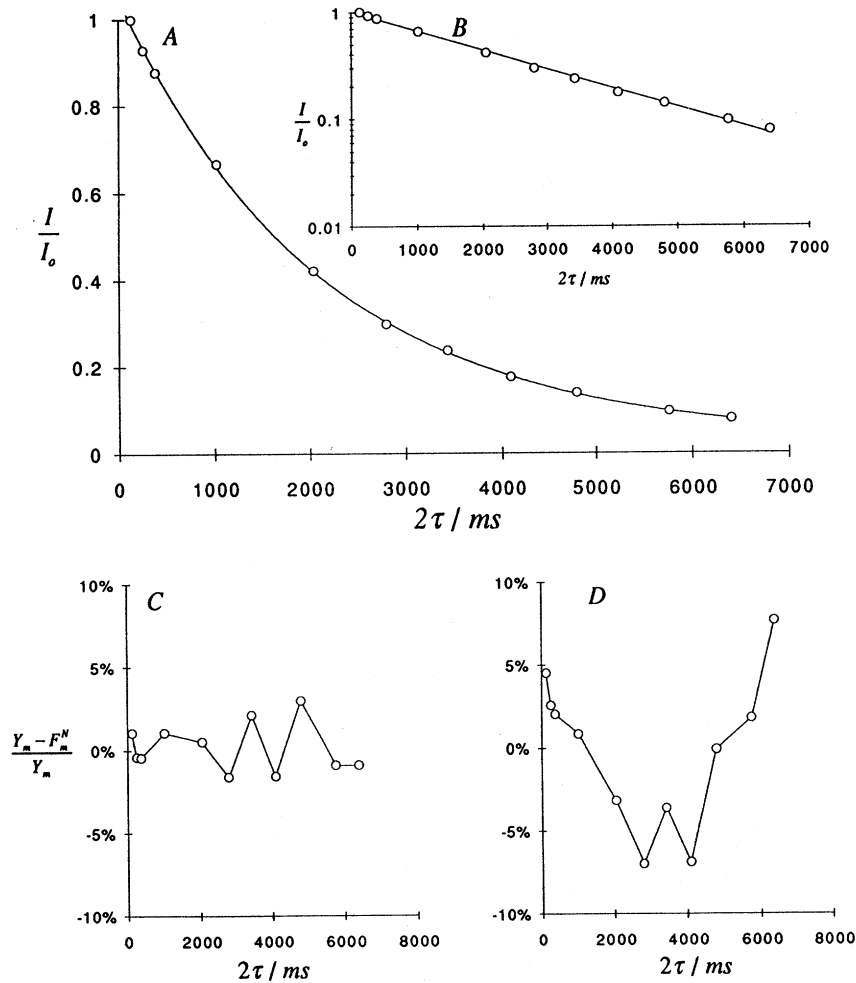


Figure 4. Carr-Purcell-Meiboom-Gill (Eq. [15]) data plotted using the G-N fit (A) and the semilog plot (B). Deviations in $(Y_m - F_m^n)/Y_m$ as a function of 2τ for (A) and (B) are shown in (C) and (D), respectively.

least five or more T_1 values to estimate I_∞ (12). Because most measurements at long τ values cannot be used in the semilog linear form, the standard error terms increase with vanishing degrees of freedom ($df = M - L$). Looking at the T_2 data (Fig. 4A and 4B), we see that the G-N method provides a better fit to the observed data, as evidenced by smaller deviations in $Y_m - F_m^N$ (Fig. 4C compared with 4D).

Of course, linear regression methods are not always applicable; many useful empirical relationships (Table 1), including our example discussed above, are not readily transformed for linear regression analysis. Another such expression is shown in Fig. 5, where we have measured the intensity of a resonance as the pulse width was increased (phased when $PW \ll 90^\circ$) and fit the data to the trigonometric function (Eq. [16]). Although such an experiment is not a standard procedure for establishing a 90° pulse, it is a useful check of probe performance. Other uses for G-N curve fitting in NMR are exemplified by deconvolution (Eq. [17] and [18]). Of course, for these procedures our template is not useful for fitting more than two or three Gaussian or Lorentzian (or any combination thereof) lines to partially resolved resonances because of computational time limitations.

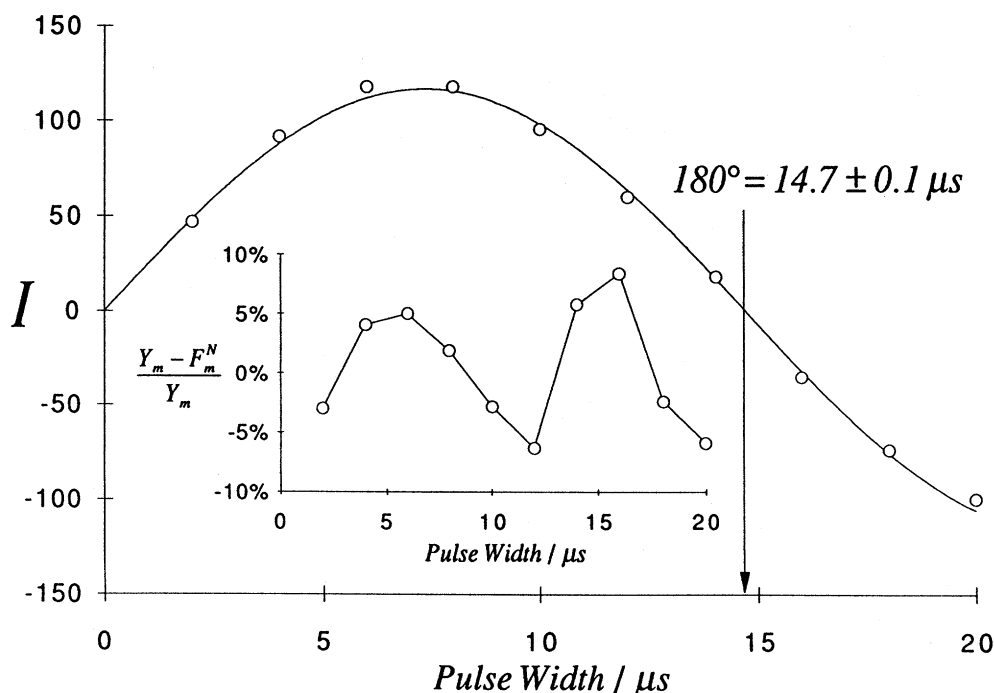


Figure 5. Changes in resonance intensity, I , as a function of pulse width (Eq. [16]) using the G-N best fit. Deviations in $(Y_m - F_m^N)/Y_m$ as a function of PW are shown in the inset.

CONCLUSION

In this article we have described a simple G-N linearization procedure using a spreadsheet. We have shown that such curve-fitting techniques are statistically more appropriate than linear regression analysis of mathematically transformed data. We have used this template to analyze data from numerous experiments (14-17) but have found that it is cumbersome when the data or parameter arrays are large (e.g., $M > 50$, $L \geq 5$).

REFERENCES

1. H. O. Hartley, "The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares," *Technometrics*, 1961, 3, 269-280.

Curve Fitting in Nuclear Magnetic Resonance Spectroscopy

2. T. C. Farrar, *An Introduction to Pulse NMR Spectroscopy*, Farragut Press, Chicago, 1987, pp. 47-48.
3. R. G. D. Steel and J. H. Torrie, *Principles and Procedures of Statistics*, McGraw-Hill, New York, 1960, p. 170.
4. C. R. Fleenor, M. E. Shanks, and C. F. Brumfiel, *The Elementary Functions*, Addison-Wesley, Reading, Massachusetts, 1968, p. 8.
5. N. R. Draper and H. Smith, *Applied Regression Analysis*, 2nd ed., Wiley, New York, 1980, pp. 458-529.
6. A. A. Afifi and S. P. Azen, *Statistical Analyses: A Computer-Oriented Approach*, 2nd ed., Academic Press, New York, 1979, pp. 182-190.
7. *Microsoft Excel User's Guide 2* (Version 4.0, Apple Macintosh Series), Microsoft Corporation, Redmond, Washington, 1992, pp. 13-14.
8. J. E. Wertz and J. R. Bolton, *Electron Spin Resonance*, Chapman and Hall, New York, 1986, pp. 396-404.
9. H. F. Davis, *Introduction to Vector Analysis*, Allyn and Bacon, Boston, 1973, pp. 216-247.
10. J. Schaefer and E. O. Stejskal, "High-Resolution ^{13}C NMR of Solid Polymers," *Topics in C-13 NMR Spectroscopy*, 1979, 3, 283-324.
11. S. Wolfram, *Mathematica, A System for Doing Mathematics by Computer*, 2nd ed., Addison-Wesley, Redwood City, California, 1991, pp. 624-630.
12. J. S. Frye, "Comparison of Inversion-Recovery Methods for Measuring Longitudinal Relaxation Rates," *Concepts Magn. Reson.*, 1989, 1, 27-33.
13. D. D. Traficante, "Relaxation — Can T_2 Be Longer Than T_1 ?" *Concepts Magn. Reson.*, 1991, 3, 171-177.
14. L. W. Doner and P. L. Irwin, "Assay of Reducing End-Groups in Oligosaccharide Homologues with 2,2'-Bicinchoninate," *Anal. Biochem.*, 1992, 202, 50-53.
15. P. Irwin, P. Pfeffer, L. Doner, and J. Ferretti, "NMR Studies of H_2O :*N*-Phenyl-(*N*-Phenyl- β -D-Glucopyranosylamine) Uronamide Interactions in Dimethyl Sulfoxide at Two Fields," *J. Carbohydr. Chem.*, 1993, 12, 63-79.
16. S. F. Osman, W. F. Fett, P. L. Irwin, D. G. Bailey, N. Parris, and J. V. O'Connor, "Isolation and Characterization of an Exopolysaccharide Depolymerase from *Pseudomonas marginalis* HTO41B," *Curr. Microbiol.*, 1993, 26, 299-304.
17. L. W. Doner, G. Bécard, and P. L. Irwin, "Binding of Flavonoids by Polyvinyl-polypyrrolidone," *J. Agric. Food Chem.*, 1993, 41, 753-757.